IN THE CLAIMS

Please **amend** the claims as indicated:

1-7. (cancelled)

8. (currently amended)   An information search method for crawling a web site via a network using a computer, said method comprising the steps of:

acquiring a web page as initial information and storing source code into a storage device;

reading the source code of said web page from said storage device;

conducting a structure analysis of said web page, wherein the structure analysis includes the steps of:

reading an HTML document of a web page as an analyzing object,

conducting a temporary block analysis based on a description of HTML tags of the HTML document,

using the HTML tags to temporarily divide the HTML document into blocks, and

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT IMAGE describes a type of media used to display the HTML document;
[[and]]

storing a result of the analysis into said storing device;

calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device; and

accessing the web site depending on the calculated degree of significance to acquire contents thereof, and storing them into said storage device.

9. (cancelled)

10. (currently amended)   A program product for controlling a computer connected to a network so as to crawl a web site, said program product causing said computer to execute:

a process of acquiring a web page as initial information and storing source code into a

JP920020109US1 — Amendment A   -2-   10/621,474

storage device;

a process of reading the source code of said web page from said storage device, conducting a structure analysis of said web page, and storing a result of the analysis into said storing device;

a process of calculating a degree of significance of a web site linking from said web page, based on the result of said structure analysis stored in said storage device, wherein scores used to calculate a degree of significance are calculated based on information elements added to anchors through an analysis of the web page, wherein the analysis of the web page includes the steps of:

reading an HTML document of a web page as an analyzing object,

conducting a temporary block analysis based on a description of HTML tags of the HTML document,

using the HTML tags to temporarily divide the HTML document into blocks, and

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document; and

a process of accessing the web site depending on the calculated degree of significance to acquire contents thereof, and storing them into said storage device.

11. (original) A program product according to Claim 10, wherein said program product causes said computer to conduct said structure analysis by associating mutually relevant information elements with each other, among information elements contained in said source code.

12. (original) A program product according to Claim 10, wherein, in the process of calculating the degree of significance of said web site, plural strategies are used as strategies for calculating the degree of significance of said web site, by giving weights thereto, respectively.

13. (currently amended)   A program product for controlling a computer so as to analyze an HTML document structure, said program product causing said computer to execute:

a first process of reading an HTML document being a processing object from a memory, blocking information elements forming said HTML document based on tags of said HTML

document, and storing blocked structural data of said HTML document into the memory; and

a second process of reading the blocked structural data of said HTML document from said memory, updating block structures of said HTML document by associating the information elements that are mutually relevant in terms of a meaning, and storing the updated structural data into the memory, wherein said second process includes the steps of:

identifying an unnecessary information element in terms [[of a purpose]] of a document structure analysis, wherein an information element is deemed to be unnecessary if the information element includes an OBJECT_IMAGE that includes a Uniform Resource Locator (URL) that has been used by another information element in the HTML document, wherein the OBJECT_IMAGE describes a type of media used to display the HTML document;

merging said information elements or dividing a block based on contents of said information elements; and

merging the block structures based on information contained in each block.

14. (cancelled)

15. (new)   A method comprising:

reading an HTML document of a web page as an analyzing object;

conducting a temporary block analysis based on a description of HTML tags of the HTML document;

using the HTML tags to temporarily divide the HTML document into blocks;

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements include plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document;

deleting any block in the HTML document that is deemed to be structurally meaningless, wherein a block is deemed to be structurally meaningless if that block has only unnecessary information elements; and

merging relevant information elements in a same block into one composite element.

16. (new)   The method of claim 15, wherein the unnecessary information elements include

JP920020109US1 – Amendment A   -4-   10/621,474

OBJECT_ANCHORS that have a same title, wherein an OBJECT_ANCHOR describes a correlation between the HTML document and elements in another web page.

17. (new)   The method of claim 16, wherein the unnecessary information elements include OBJECT_TEXT_BLOCKS that have a same description of text in a block.

18. (new)   The method of claim 17, wherein the relevant information elements that are merged are from a group that includes the OBJECT_IMAGE, OBJECT_ANCHOR and OBJECT_TEXT_BLOCKS.

19. (new)   A method for eliminating ambiguity of a specified topic being searched during a web crawling, the method comprising:

presenting relevant keywords to a user during web crawling, wherein the relevant keywords describe multiple attributes of a term that has an ambiguous meaning, and wherein the user is afforded an ability to specify keywords that have a minus degree of significance to a meaning intended by the user for web crawling; and

narrowing down crawling objects by eliminating user-specified keywords that have a minus degree of significance, thereby eliminating ambiguity of a term being searched.

20. (new)   A web crawler comprising:

an initial site acquiring section, wherein the initial site acquiring section specifies a Uniform Resource Locator (URL) of a home page of a specific web site from which information is to be collected, and wherein initial web sites to be searched are obtained through the use of keywords in a search engine, and wherein the initial web sites represent a set of web sites that are initially set for web crawling;

a document structure analysis section for performing document structure analysis for a web page of the initial sites, wherein the document structure analysis includes the steps of:

reading an HTML document of a web page as an analyzing object,

conducting a temporary block analysis based on a description of HTML tags of the HTML document,

using the HTML tags to temporarily divide the HTML document into blocks, and

identifying unnecessary information elements in the HTML document, wherein the unnecessary information elements are plural information elements that include an OBJECT_IMAGE having a same Uniform Resource Locator (URL), wherein the OBJECT_IMAGE describes a type of media used to display the HTML document;

a significance calculating section for calculating degrees of significance of web sites that are acquired by web crawling, wherein the degrees of significance are based on a result of the document structure analysis performed by the document structure analysis section, and wherein calculating degrees of significance extends a Fish-Search crawling technique by basing the calculating on strategies specified by a user and information elements added to anchors through the document structure analysis performed by the document structure analysis section, and wherein objects of crawling are dynamically determined depending on the degrees of significance; and

a crawling executing section for executing a process of acquiring web sites by crawling based on the results of the degrees of significance calculated by the significance calculating section.